

Content Management, Portals, and Personalization

*Phillip J. Windley, Ph.D.
Chief Information Officer
Office of the Governor
State of Utah*



Maintaining a web site seems pretty simple at first blush. After all, the concepts are fairly straightforward and even if you don't want to learn HTML, there are plenty of editing tools that ensure you can build nice web pages without bothering yourself with the technical details. After that, all you need is put your pages on a web server and you've published a web site.

Things begin to change however when those few pages you started out with blossom to tens, hundreds, or even thousands of pages with some dynamic content. At that point even simple changes like updating a link can be so much work that they don't get done. A larger change like updating the look and feel can take months and cause people to quit. Compounding the problem, as the web site becomes more complicated to maintain, it gets turned over to technical personnel who can't possibly maintain pages that require subject matter expertise.

Eventually the problem deteriorates to an endless dance between the technical folks and the business managers trying to coordinate changes to a web site where they both need to be present for even the simplest change. Fortunately this needn't be the case. Content management offers a solution to managing large web sites with multiple content owners and keeping your sanity.

Web Site Basics

All web sites have three things in common:

1. Content – the content is the meat of the web site. People come to web sites for their content. Content is usually produced and maintained by a subject matter expert who we'll call the "custodian."
2. Layout – the layout is the format that the content takes on the web site. Choices about fonts, colors, headers, footers, and the position of links are all layout choices.
3. Navigation – navigation is the part of the web site that let's you go from one page to another. In some cases navigation can be very rich, allowing a user multiple choices. In other cases, its might be more restricted to guide a user through a particular set of content.

Usually, layout and navigation decisions are best made by someone with a talent, or better yet training, in user interface design and implemented by technicians who understand the intricacy over HTML, cascading style sheets, and programming.

Separating concerns about these web site components allows us to manage a large web site in a much more sane and rational way. Software tools, called *content managers*, combine content, descriptions of the navigation, and layout templates to produce a completed web site. Once the separation is made and tool for reconstituting the final web site has been chosen, the content files can be maintained by the subject matter expert and the layout and navigation are maintained by the technical experts. Although choices about what the layout looks like and how navigation works are still business choices, their implementation is much more technical and can be done according to specification.

Even more importantly, the business expert can maintain the content without the distractions caused by navigation and layout. In a site consisting of thousands of pages, each page could have a different owner who maintains the content according to their individual expertise and yet the entire site could have a common navigation structure and a common look and feel.

Flexible Presentation

Content management not only reduces the workload of building and maintaining a large web site, but also enables a number of other important and beneficial features. Once the content has been separated from the layout and site navigation, it can be used in many different ways besides a single presentation on a web site. This flexibility in presentation allows the content to be used in multiple arenas.

Once upon a time, a single browser, Mosaic, was all anyone had and all that site designers had to design web sites to operate on. Today, however, not only are there multiple browsers, but each of those browsers have multiple versions still in use. Compounding those issues are the fact that the browsers now run on multiple platforms with widely varying form factors, such as PDAs, phones, and pagers. Further, a web site designer needs to be concerned with accessibility issues to be in compliance with ADA and Section 508 requirements.

Doing all of this by hand is impossible. With a content management system, however, the problem is solved once, in a series of templates, and then the entire site will be accessible from multiple browsers on multiple platforms, and maintain accessibility standards.

An even more interesting use of the ability to have flexible presentation for content is the presentation of content for use by other programs (as opposed to presenting it for use by a human in a browser). The content could be turned into

any format imaginable, but the most common format would be some kind of XML. An interesting use of this ability is called “content syndication.”

Content Syndication

Content syndication, also called “publish and subscribe” is a way for content, independent of its layout, to be published to subscribers who then combine it with other information and display it for their own use.

Portals, like Yahoo!, are examples of subscribers to content published by someone else. For example, my Yahoo! page shows me the headlines from *USA Today* right next to the financial news from *Bloomberg*. Yahoo! isn’t producing the content, *USA Today* and *Bloomberg* are. Yahoo! is subscribing to their *syndicated* content and then presenting it to me in a format completely different from what I’d see if I went to the *USA Today* or *Bloomberg* homepages directly. Please note that this content is dynamic. Each time I go to Yahoo!, the information has changed based on what *USA Today* and *Bloomberg* have done in managing their content.

There are ways to make this sort of publish and subscribe mechanism much more personal. An XML protocol called RSS (Rich Site Summary) enables web content to be summarized and placed on a web server. People who wish to subscribe to that content and view it using programs called

“aggregators” to retrieve multiple RSS feeds and then combine, render, and view them. The difference between doing this and using a web browser is that I can view summaries of a number of different web sites all in one place without having to visit each one. Further, I am deciding what content I see, not Yahoo! RSS feeds are available for a number of sites, including the New York Times.

Typically people refer to these feeds as “news sources” but its important to consider that “news sources” could also mean reports generated by your accounting software, status of your servers, posts in a discussion group, orders from your e-commerce site, updates from your co-workers workflow management software, and so forth.

We can make use of content syndication for state data in exactly this same way. By providing RSS and other XML based feeds of important data, we enable it to

Resources

The following resources will help you learn more about content management, syndication, and portals.

[Interwoven Content Management Suite](#) – The Interwoven content management suite is a set of tools for managing large, complex web sites with multiple content users. The state has selected Interwoven as the content manager for the state portal.

[Novell Portal product](#) – Novell’s portal tool allows for personalization and works well with the state’s directory and Interwoven.

[Metabrowser Tutorial](#) – The state library is actively working on using meta data to catalogue state web pages. This tutorial shows how to use a metabrowser to put Dublin Core metadata in your web pages.

[Training and Tools for Metadata](#) – The GILS project provides training and tools for metatagging your content.

[RSS Tutorial](#) – The folks at GILS (part of the state library) have written an excellent tutorial on metadata and RSS.

[RSS](#) – RSS is the standard for syndicating content. RSS is based on XML.

[Aggregators](#) – There are number of RSS aggregators available. Most of them are fairly early programs, much like early browsers, but show the potential of this new class of tools.

[Radio Userland](#) – If you’d like to experiment with content management and syndication, you can find out more using Radio.

be shared and used both across agency boundaries and by our clients¹. As an example, suppose that the courts published opinions and verdicts as an RSS feed. A lawyer could subscribe to the feed for the court of interest and be notified on her desktop when opinions and verdicts came out in real time. We all provide data to others that would be potentially more useful formatted as an RSS feed or with some other XML standard.

A content management system makes such syndication possible and largely automatic once its set up. If we're maintaining content in web pages, providing an RSS feed is an arduous editing process. If we're maintaining our content in a content management system, then the program can automatically render the content as web pages for humans, as RSS feeds for content aggregators, and in any other format we need.

Metadata

Web pages usually contain more content than just what you see on your screen. Almost every web page contains a header that gives the title of the page. If the page was created by an HTML editor or other program, the page will also usually contain a tag giving the name and version of the program. Web page headers can also contain other data that describes the page; such data is called *metadata* because its data describing the data in the web page. Metadata might tell when the page was created, who the author was, give keywords or other information useful for searching on the page, link data, and so on.

The State Library runs a program called the Government Information Locator Service (GILS) which researches and proposes techniques that Utah government can make its online information more easily retrievable. They are nationally recognized for their work.

The GILS project and the CIO's office endorse a metadata standard called the Dublin Core. They have developed templates and schemas based on the Dublin Core and specific to Utah State government. Incorporating these elements in a web page makes its easier for search engines to correctly categorize your document and makes it more likely that it will be found.

Personalization

Personalization uses information about the viewer of a web page to change the content, layout, and navigation of that page to better suit the viewer's preferences or needs. If you've used my.yahoo.com or any other large portal, you've probably been exposed to personalization.

In its simplest form, personalization merely changes some of the least important content in response to information the web site knows about you (see my earlier

¹ Note that we're not dealing with authentication and authorization issues in this document. The same authentication and authorization needs to be done for syndicated content as for web pages to ensure that only those who are authorized to get a set of data have access.

paper on Cookies and Privacy for more information on *how* they know who you are). For example, a web site might say “Hello Phil” at the top of the page without changing the real content of the site. The layout, colors, and fonts of a site are also relatively easy thing to personalize with a content management system since the page can be rendered in response to the user’s preferences.

More sophisticated personalization schemes change the content and even the navigation in response to user preferences. For example, if I might chose to see CommuterLink information and renewal notices from Commerce on my homepage. Someone who doesn’t commute or have a corporation wouldn’t care about this content. Notice that in the first instance, the personalized content is relatively generic—I’d see the same CommuterLink map regardless of who I am. In the second case, the personalization is specific to me and my interaction with the Department of Commerce.

Recently Utah state government consolidated all of our authentication and authorization information for state employees into a single master authentication directory. This makes personalization on InnerWeb possible. Your personal Innerweb page could show you news from Capitol Connections, filtered to your interests, your retirement information, and your annual leave balance.

In the same way, we can store citizen personalization profiles in the same master authentication to create personalized utah.gov portals for citizens. The use of personalization will be optional (opt-in) and recent changes to the GRAMA statute protect this database as private.

Portals

Once we have content that is separate from its layout and navigation, we have made it available through syndication, and we can keep track of user preferences, a software tool called a portal can be used to render personalized content dynamically changing the content, the layout, and the navigation in response to not only the user’s personal preferences, but their mode of access and other factors.

The portal software combines user profile information, content that is syndicated by agencies and others, along with information about the current connection to render pages specific to that user. Once we put all this together, I could use utah.gov from my web browser, from a content aggregator, from some other specialized application, from my PDA, or my mobile phone and get content that is specific to me, formatted for the device I’m using.

From Content to Transactions

So far, we’ve been talking about content in terms of information or data. That begs the questions: what about transactions? If content management and portal software provide a way to easily share and aggregate data from multiple sources, can we do the same thing for transactions?

The answer to this question is “yes” and the methods for doing it is commonly called “web services.” In fact, RSS and other content syndication mechanisms are simple forms of web services. Just as building a transactional web site is more complicated than building a web site consisting of nothing but static content, building a transactional web site from multiple syndicated services is more complicated than aggregating data and producing personalized pages. Still, the technology to do it is becoming available. I’ll discuss web services in a future paper.

Conclusions

Content management software offers us the ability to apply the right skill sets to the right tasks and reduce our workloads at the same time. Portal software allows us to use this structure to provide our users with better service by putting the information they care about on the utah.gov. Utah has bought the Interwoven content management suite and Novell’s portal software. They should be ready for use in Fall 2002. Now would be a good time to start planning for these new features and asking what data your constituents would like to see you syndicate.